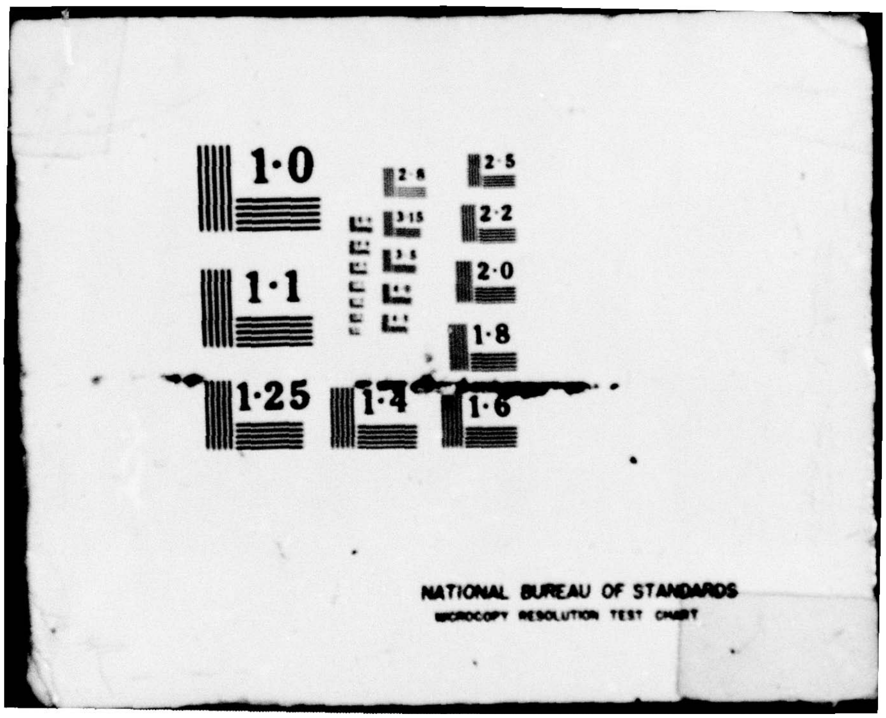
AD-A078 510 CALIFORNIA UNIV LOS ANGELES DEPT OF SYSTEM SCIENCE F/G 12/1
OPTIMAL INVESTIGATION AS A REGENERATIVE STOPPING PROBLEM.(U)
MAY 79 A G BUCKMAN , B L MILLER N00014-78-C-0428
UNCLASSIFIED UCLA-WP-289 NL

1 OF 1
AD
A 078510

END
DATE
FILMED
1 -80
DDC


1·0
2·8
2·5
3·15
2·2
2·0
1·1
1·8
1·25
1·4
1·6
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

Working Paper No. 289

OPTIMAL INVESTIGATION AS A REGENERATIVE STOPPING PROBLEM

by

A. GREGORY BUCKMAN
School of Business Administration
California State University, Hayward

and

BRUCE L. MILLER
Department of System Science
University of California, Los Angeles

May 1979

WESTERN MANAGEMENT SCIENCE INSTITUTE
University of [redacted] Los Angeles

Working Paper No. 289

OPTIMAL INVESTIGATION AS A REGENERATIVE STOPPING PROBLEM*

by

A. Gregory Buckman
School of Business Administration
California State University, Hayward

UCLA-WP-289

and

Bruce L. Miller
Department of System Science
University of California, Los Angeles

May 1979

25

*We thank K. S. Azoury and I. Costea for helpful comments on an earlier draft of this paper. This research was partially supported by the National Science Foundation under Contract ENG 76-12250 A01 and the Office of Naval Research under Contract N00014-78-C-0428.

NSF-ENG76-12250

405 213

ABSTRACT

THIS REPORT DISCUSS ~~We reconsider~~ the problem of dynamic optimal investigation of a two-state (in control, out of control) system. The true state can only be inferred from reported costs and the time since the last correction. It is demonstrated that when the parameters satisfy certain conditions, such problems can be efficiently solved as regenerative stopping problems. Some general results for regenerative stopping problems are also obtained. In the last section the problem is generalized to n two-state systems. By combining simulation with the Regenerative Stopping Algorithm, a problem with 20 state variables is solved with a small error term.

1. INTRODUCTION

Management is often faced with the problem of assessing financial reports to determine if corrective action should be taken. The situation is probabilistic in that disappointing performance does not necessarily mean that corrective action should be taken since the disappointing performance may be caused by some uncontrollable and nonrecurring factor. This problem area has attracted the attention of Buckman [6], Dittman and Prakash [9] Dyckman [10], Ozan and Dyckman [15], and Magee [13]. In [12] Kaplan has a survey of this area which includes some industrial engineering models. Mathematically these problems are similar to some of the maintenance models, and the reader is directed to Pierskalla and Voelker [16] for a survey of the maintenance model literature.

The approach of Kaplan [11] is dynamic and allows the decisions each period to depend on our current estimate of the probability that the system is in control. Since the true state of the world is not known, we have a difficult optimization problem which would seem to permit only the two-state (in control, out of control) model considered by Kaplan, and even this model can be very demanding computationally. Magee [13] considers Kaplan's two-state model, and allows costs to be normally distributed so that the true state can take on all values on subsets of [0,1]. Because of the difficulties of computing an optimal rule, Magee proposes seven plausible rules which he compares by simulation. The non-dynamic formulation of Ozan and Dyckman [15] allows a much more detailed model which is solved by linear programming rather than dynamic programming. For example, cost variances are permitted to have a number of possible causes.

We consider the two-state Kaplan model and solve it as a regenerative stopping problem. However unlike Kaplan's value iteration solution procedure,

or the policy iteration procedures mentioned below, we must require that the parameters of the model satisfy certain restrictions. Regenerative stopping problems were formulated as such independently by Brender [4] and Breiman [3]. Algorithms for such problems are not fully developed, and this paper gives some new results. Besides being more efficient than Kaplan's algorithm, our procedure is evidently more efficient than the policy iteration methods of Brown [5] and Sondik [19] which can also be applied to Kaplan's model.

In the last section we generalize the Kaplan model to n dimensions by considering n independent processes which are related in that corrective action is taken for all n for none of them. As n increases the problem rapidly becomes intractable for all procedures including ours. However, by using simulation techniques, the n state variable problem can be solved by the regenerative stopping algorithm. This introduces an error factor but 20 state variable problems are solved, and policies are obtained which come within .00003 percent of the true minimum cost.

2. THE KAPLAN MODEL

We consider a two-state production system where state 1 means the system is "in control" and state 2 means the system is "not in control." We let Y_i be the random variable which represents the reported costs in period i. When the system is in state k, k = 1,2, the probability mass function of reported costs is $f_k(y)$. Presumably $f_1(y)$ has most of its probability at low costs and $f_2(y)$ has most of its probability at higher costs. We let m_1 and m_2 be the means of the two distributions.

When in state 1 there is a probability p of staying in state 1 and probability (1-p) of going to state 2. This move to state 2 is assumed to take place late enough so that reported costs are determined by the state at the

beginning of the period and therefore are described by f_1[†]. The cost in the given period and the state the system goes to in the next period are assumed to be conditionally independent given the state at the beginning of the period. When in state 2 the system will stay there if no corrective action is taken. Therefore, the system can be represented by a two-state Markov process whose one-step transition matrix is:

$$P = \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix}$$

Investigation and correction may be taken at the beginning of any period at a cost K and the system goes to state 1 instantaneously. The cost K is incurred even if the system were already in state 1 and there was no correction.

The state of the system is not known to the decision-maker except after an investigation, and it can only be inferred from the operating costs and the length of time since corrective action. What is actually determined is the probability that the system is in state 1. This is carried out by the following Bayesian procedure. Let x_i be the probability that the system is in state 1 at the beginning of period i before knowing the reported costs. During period i the reported costs are $Y_i = y$, and after obtaining the reported costs we make a revised estimate of the probability that the system is in state 1 at the beginning of period i, $\hat{x}_i$, and

$$\hat{x}_i = \frac{f_1(y)x_i}{f_1(y)x_i + f_2(y)(1-x_i)} \qquad (1)$$

[†] Kaplan's position on this issue is unclear. On Page 33 he says that the move is early enough in the period to effect operating costs. On Pages 35 and 36 costs are based on the ending state immediately following an investigation, but are based on the beginning state otherwise. The latter is consistent with the Bayesian equation on Page 34, which holds only if costs are based on the beginning state.

As stated earlier equation (1) assumes that if there is a transition from state 1 to state 2 the costs are determined by the state at the beginning of the period, state 1. The value of x_{i+1} is then

$$x_{i+1} = p\hat{x}_i + 0(1-\hat{x}_i) = p\hat{x}_i , \tag{2}$$

which is the same as the equation of Page 34 of Kaplan [11].

Since the true state cannot be observed directly, the state variable of our Markov decision problem is x_i. Two decisions are available, decision 1 which is do nothing, and decision 2 which is to investigate and correct if necessary. The one period expected costs are given by

$$C(x_i,1) = x_i m_1 + (1-x_i)m_2 \tag{3}$$

$$C(x_i,2) = K + m_1 \tag{4}$$

Let d_i be the decision in state i. Then the state transition function is given by

$$x_{i+1} = p\hat{x}_i \text{ if } d_i = 1 \text{ and } Y_i = y \text{ and} \tag{5}$$

$$= p \text{ if } d_i = 2.$$

The probability that $Y_i = y$ is $f_1(y)x_i + f_2(y)(1-x_i)$.

The objective is to determine a decision rule (a specified decision for each possible state and period) which minimizes the expected discounted cost over an infinite planning horizon. The discount factor α, $0 < \alpha < 1$, will represent the time value of money to the firm.

Kaplan solved his problem by solving for F_1, the optimal return function with one period to go, for all $0 \leq x \leq 1.0$, then for F_2, the optimal return function with two periods to go, and then for F_3, etc. The convergence of the F_i is infinite and in value iteration the repetition of a policy (the same decision rule for F_i as F_{i+1}) does not imply that the policy is optimal. Although

these difficulties are not discussed by Kaplan, presumably one could use the ideas found in Porteus [17] to determine when an optimal policy has been found.

It is interesting that Kaplan's problem can also be solved by policy iteration, and the methods are finite for Kaplan's problem when the number of realized costs is finite. Either Brown's method of recursive sets of rules [5] on Sondik's "finitely transient" procedure [19] can be used, and Sondik points out their similarity. The computational requirements of Brown's method when applied to an example problem is discussed in Section 4.

3. REGENERATIVE STOPPING PROBLEMS

We propose to solve the Kaplan model as a regenerative stopping problem. Since regenerative stopping problems are optimal stopping problems which regenerate or recommence upon stopping, we begin by describing optimal stopping problems.

The general optimal stopping problem is described at the beginning of Chapter 3 of Chow, Robbins, and Siegmund [7]. We will be less general and consider a stopping problem which is a Markov decision problem with the set of possible states X. In the Kaplan model X is the interval [0,1]. When the system is in state $x \in X$, one decision is to continue and receive a cost $C(x,1)$ and go to the next state according to the probability mass functions p_{xz}, $z \in T_x$, where T_x is the countable number of new states which can be reached from x. The other decision is to stop and receive a cost of $C(x,2)$. The problem ends immediately with the decision stop. Costs are discounted by a discount factor α. Our objective is to determine a decision rule which minimizes the expected discounted cost incurred up to and including stopping where the initial state is x^o. We assume that the costs $C(x,1)$ and $C(x,2)$ are bounded in absolute value as x varies over X.

In order for the optimal stopping problem to have an easily computable solution it is necessary that it satisfy the Monotone Condition.

Monotone Condition. Let $B = \{x : C(x,2) \le C(x,1) + \alpha \sum_{z \in T_x} C(z,2) p_{xz}\}$. Then $\sum_{z \in T_x \cap B} p_{xz} = 1$ if $x \in B$.

The set B represents precisely those states where stopping is at least as good as continuing exactly one more period and then stopping. In problems with enough structure that the monotone condition holds, the set B is often very easy to determine. The monotone condition says that when the system reaches the set B it stays in B.

Clearly if $x \notin B$, then the optimal decision is to continue. It also would seem to be true that if $x \in B$ and the monotone condition holds then the optimal decision is to stop. This needs to be proven and is known as the Monotone Stopping Theorem (Chow, Robbins, and Siegmund [7, Theorem 3.3], Ross [18,Theorem 6.14)]). In addition to the monotone condition, some additional technical conditions are needed to prove the theorem, which vary with the particular form of the optimal stopping problem considered. Ross's proof [18], Theorem 6.14] applies to our formulation once his stability condition is verified. The stability condition is that G_n, the optimal return function when there are n periods to go, converges to G, the optimal return function for the infinite horizon case. We have assumed that C(x,1) and C(x,2) are bounded so that it is well-known (Denardo [8, Theorem 4]) that $G_n \to G$, and therefore the Monotone Stopping Theorem does apply to our formulation with bounded costs.

Theorem 1. (The Monotone Stopping Theorem) Suppose that the monotone condition holds. Then the optimal decision rule is to continue if $x \notin B$ and to stop if $x \in B$.

The Monotone Stopping Theorem is important computationally since the

optimal policy is known once B is determined, and there is no need for the recursive dynamic programming calculations. It is an example of a myopic policy being optimal, and the simplification derived from myopic policies is well-known.

Regenerative stopping problems differ from optimal stopping problems in that when we stop and receive a cost of $C(x,2)$, the system moves instantaneously to the initial state x^o, and the problem continues. Breiman [3] called such processed, "Binary Decision Renewal Problems." Our criterion will be to minimize the expected discounted cost over an infinite horizon. Solving such problems and in particular the Kaplan model is the topic of the next section.

4. A COMPUTATIONAL PROCEDURE

Regenerative stopping problems were first formulated independently by Brender [4] and Breiman [3], and both showed that regenerative stopping problems could be solved by solving the right stopping problem. They considered the nondiscounted case with the average cost per period criterion. For the discount case, this basic theorem is stated in Bell [1, Theorem 1] who attributes it to Taylor [20]. However, Theorem 4 of [20] is for the average cost criterion, and therefore a proof for the discount case will be supplied. This result suggests the strategy of solving regenerative stopping problems by solving a sequence of stopping problems ending with the right stopping problem, but neither Brender nor Breiman considered this idea.

To state the theorem we need the idea of λ-stopping problem. A λ-stopping problem is a stopping problem where the cost of continuing is changed from $C(x,1)$ to $C(x,1) - \lambda$ and the cost of stopping $C(x,2)$ is left unchanged. In the discount case Theorem 2 gives us the interpretation of λ^*, the "right" λ, as $(1-\alpha)\ F^\alpha(x^o)$ where $F^\alpha(\cdot)$ is the optimal return function of the regenerative stopping problem with a discount factor of α. In the average cost per period

case, λ stand for the average cost per period. In the finite state model, it is known (Blackwell [2,Theorem 4]) that $(1-\alpha)\ F^{\alpha}(\cdot)$ converges to the vector of average costs per period as $\alpha \to 1$.

Let $G(x,\lambda)$ be the optimal return function of the λ- stopping problem from the initial state x. The initial state x^{o} is important enough that we introduce the function V defined by $V(\lambda) = G(x^{o},\lambda)$.

Theorem 2. If λ^{*} satisfies $V(\lambda^{*}) = 0$, then the optimal decision rule for the λ^{*}-stopping problem is the optimal decision rule for the regenerative stopping problem. Also $\lambda^{*} = (1-\alpha)\ F^{\alpha}(x^{o})$.

Proof. It is convenient to drop the superscript α and let $F = F^{\alpha}$ since α is fixed in what follows. The equations of optimality are

$$G(x,\lambda^{*}) = \min\ (C(x,1) - \lambda^{*} + \alpha \sum_{z \in T_x} G(z,\lambda^{*})p_{xz},\ C(x,2)) \text{ and} \tag{6}$$

$$F(x) = \min\ (C(x,1) + \alpha \sum_{z \in T_x} F(z)p_{xz},\ C(x,2) + F(x^{o})). \tag{7}$$

There is no discount factor before $F(x^{o})$ in equation (7) since the move to x^{o} is instantaneous with the decision stop. We want to show that $\bar{F}(x) = G(x,\lambda^{*}) + \lambda^{*}/(1-\alpha)$ satisfies (7). From (6) we have that

$$\bar{F}(x) - \lambda^{*}/(1-\alpha) = \min\ (C(x,1) - \lambda^{*} - \alpha\lambda^{*}/(1-\alpha) + \alpha \sum_{z \in T_x} \bar{F}(z)p_{xz},\ C(x,2)).$$

By cancelling and adding $\lambda^{*}/(1-\alpha)$ to all terms we have

$$\bar{F}(x) = \min\ (C(x,1) + \alpha \sum_{z \in T_x} \bar{F}(z)p_{xz},\ C(x,2) + \lambda^{*}/(1-\alpha)).$$

By hypothesis $G(x^{o},\lambda^{*}) = 0$ so that $\bar{F}(x^{o}) = \lambda^{*}/(1-\alpha)$ and $\bar{F}$ satisfies (7). By the uniqueness of the solution of the equation of optimality in the discount case $\bar{F} = F$ and $\lambda^{*} = (1-\alpha)\ F(x^{o})$. Furthermore the two equations below (7) establish that the same decisions which optimize (6) for each state x optimize

(7), which completes the proof.

Thus the regenerative stopping problem can be solved by solving the simpler stopping problems if the right λ is used. Furthermore, if the monotone condition holds for the λ-stopping problem, then the Monotone Stopping Theorem can be applied. Thus, our computational approach is to solve a sequence of $V(\lambda)$ problems until a λ^* satisfying $V(\lambda^*) = 0$ is found.

We begin by observing that $V_\pi(\lambda)$, the expected cost of the λ-stopping problem using a policy π, is

$$V_\pi(\lambda) = E[C_T] - \lambda E[1-\alpha^T]/(1-\alpha) \tag{8}$$

where T is the random period where we choose to stop and C_T is the original (without subtracting λ) discounted cost up and including the cost of stopping using the policy π. Our convention is that the periods are numbered starting from zero, so that if T is the random period we stop then we have gone T periods until stopping. The interpretation of $E[1-\alpha^T]/(1-\alpha) = 1 + \alpha +..+ \alpha^{T-1}$, $T \geq 1$ $= 0$ for $T = 0$, is the expected discounted number of periods until stopping.

The following proposition is needed for the regenerative stopping algorithm. The proof is not given since it is similar to that for the analogous result in the average cost case which is proven in [14].

<u>Proposition 1</u>. V is a decreasing, finite-valued, and concave function of λ. Since V is concave it is known that the right and left hand derivative exists everywhere for $\lambda > 0$. Furthermore

$$V'_-(\lambda) \geq - E[1-\alpha^T]/(1-\alpha) \geq V'_+(\lambda)$$

where V'_+ and V'_- are the right and left hand derivatives of V, and T is the time we stop using a λ-optimal policy.

In [14] two other propositions were established which proved respectively an alternative optimality condition and showed that the solutions improve as

successive λ-stopping problems are solved. Instead of using those results we will establish a better result which gives a error bound for a non-optimal policy. The proof is given for the discount case and a similar result can be obtained for the average cost case. As a preliminary to Proposition 2, we observe that $F_\pi^\alpha(x^o)$, the expected discounted cost of the regenerative stopping problem using policy π and starting from state x^o, satisfies

$$F_\pi^\alpha(x^o) = E[C_T] + E[\alpha^T]\, F_\pi^\alpha(x^o) \tag{9}$$

where T and C_T are as defined in (8). From (8) $E[C_T] = V_\pi(\lambda) + \lambda E[1-\alpha^T]/(1-\alpha)$ and we substitute this equation into (9) and rearrange to obtain

$$(1-\alpha)\, F_\pi^\alpha(x^o) = \lambda + V_\pi(\lambda)/(E[1-\alpha^T]/(1-\alpha)). \tag{10}$$

Proposition 2. Let $\lambda_0 < \lambda_1$ be such that $V(\lambda_0) > 0$ and $V(\lambda_1) < 0$. Let π_0, π_1 and π^* be the λ_0-optimal, the λ_1-optimal and the λ^*-optimal policy of Theorem 2 respectively. Then $(1-\alpha)\,(F_{\pi_0}^\alpha(x^o) - F_{\pi^*}^\alpha(x^o)) \le \dfrac{V(\lambda_0)\;[T(\lambda_1)-T(\lambda_0)]}{T(\lambda_1)\;T(\lambda_0)}$, and

$$(1-\alpha)\,(F_{\pi_1}^\alpha(x^o) - F_{\pi^*}^\alpha(x^o)) \le \frac{-V(\lambda_1)\;[T(\lambda_1)-T(\lambda_0)]}{T(\lambda_1)\;T(\lambda_0)}$$

where $T(\lambda_0) = E[1-\alpha^T]/(1-\alpha)$ is the expected discounted time until stopping using the policy π_0 (T depends on π_0). An analogous definition applies to $T(\lambda_1)$.

Proof: By Proposition 1, V is concave and $-T(\lambda_1)$ is a supporting hyperplane at λ_1. Therefore for $\lambda_0 \le \lambda \le \lambda_1$, $V(\lambda) \ge V(\lambda_0) + (\lambda-\lambda_0)(-T(\lambda_1))$, since the one-sided derivatives of V are greater than or equal to $-T(\lambda_1)$ for $\lambda_0 \le \lambda \le \lambda_1$. Since λ^* is between λ_0 and λ_1, $0 \ge V(\lambda_0) + (\lambda^*-\lambda_0)(-T(\lambda_1))$. Dividing by $-T(\lambda_1)$ yields $0 \le V(\lambda_0)/(-T(\lambda_1)) + (\lambda^*-\lambda_0) - V(\lambda_0)/T(\lambda_0) + V(\lambda_0)/T(\lambda_0)$. We now apply (10) where $\pi = \pi_0$ and $\lambda = \lambda_0$, and $\lambda^* = (1-\alpha)\, F_{\pi^*}^\alpha(x^o)$ to obtain,

$$\frac{V(\lambda_0)}{T(\lambda_1)} - \frac{V(\lambda_0)}{T(\lambda_0)} \leq (1-\alpha)(F^{\alpha}_{\pi^*}(x^o) - F^{\alpha}_{\pi_0}(x^o))$$

which establishes the first inequality. The second is established in a similar manner starting from the equation

$$V(\lambda) \geq V(\lambda_1) + (\lambda-\lambda_1)(-T(\lambda_0)) \text{ for } \lambda_0 \leq \lambda \leq \lambda_1. \quad \text{Q.E.D.}$$

As a corollary to Proposition 2 we see that if $T(\lambda_0) = T(\lambda_1)$ then both the λ_0-optimal and λ_1-optimal policies are optimal. In the discounted case, Proposition 1 says that V is finite and decreasing so that there will always exist a λ^* such that $V(\lambda^*) = 0$. It could be the case that the expected time until stopping for the optimal policy is infinite.

<u>The Regenerative Stopping Algorithm</u>

<u>Step 0.A</u>. Find a λ_0 which is less than λ^*, where by Theorem 2 $\lambda^* = (1-\alpha)\ F^{\alpha}(x^o)$. It is desirable that λ_0 be as large as possible. We solve the λ_0-stopping problem and let π be the optimal policy for that problem. Since $\lambda_0 \leq \lambda^*$, $V(\lambda_0) \geq 0$. If a mistake is made and λ_0 is greater than λ^*, then $V(\lambda_0) < 0$ and λ_0 can be changed until $V(\lambda_0) \geq 0$.

<u>Step 0.B</u>. Set $\lambda_1 = \lambda_0 + V(\lambda_0)/(E[1-\alpha^T]/(1-\alpha) = (1-\alpha)\ F^{\alpha}_{\pi}(x^o) \geq \lambda^*$ where π is the optimal policy of the λ_0-stopping problem. We solve the λ_1-stopping problem. Since $\lambda_1 \geq \lambda^*$, $V(\lambda_1) \leq 0$.

<u>Step 1</u>. We are now in the general case where we have solved a λ_0-stopping problem and a λ_1-stopping problem where $\lambda_0 \leq \lambda^*$ and $\lambda_1 \geq \lambda^*$. The new λ-stopping problem to be solved is given by $\lambda^{new} = \min(\lambda, (1-\alpha)\ F^{\alpha}_{\pi}(x^o))$ where π is the best policy determined to date, and $\lambda = \alpha\lambda_B + (1-\alpha)\ \lambda_A$ where $0 < \alpha < 1$. The subscript B stands for bisection and the subscript A stands for approximation. Computational experiences suggest choosing a low value of α, since the approximation is quite accurate. We have $\lambda_B = 1/2\ \lambda_0 + 1/2\ \lambda_1$, and

λ_A is the λ such that $V_A(\lambda) = 0$, where $V_A(\lambda)$ is based on the four equations:

$$V_A(\lambda_0) = V(\lambda_0) \quad , \quad V'_A(\lambda_0) = -T(\lambda_0),$$
$$V_A(\lambda_1) = V(\lambda_1) \quad , \quad \text{and } V'_A(\lambda_1) = -T(\lambda_1) \tag{11}$$

where $T(\lambda_0)$ and $T(\lambda_1)$ are defined in Proposition 2. These equations determine the coefficients of the cubic approximation $V_A(\lambda) = B_0 + B_1\lambda + B_2\lambda^2 + B_3\lambda^3$. The conditions on the derivatives are based on Proposition 1. The λ^{new}-stopping problem is solved and λ^{new} replaces λ_1 if $V(\lambda^{new}) < 0$ and replaces λ_0 if $V(\lambda^{new}) > 0$. We check Proposition 2 to see if the error term is sufficiently small. If not return to Step 1.

Comment. A value of $\alpha > 0$ in Step 1 assures that the "interval of uncertanity" goes to zero. For further discussion of the Regenerative Stopping Algorithm the reader is referred to [14].

Example

We will apply our computational procecure to the investigation model of Kaplan. In this numerical example, there are only two cost outcomes, zero and six. If the state is 1 then sixty percent of the time the cost is 0 and forty percent it is 6. If the state is 2, the costs 0 and 6 are equally likely. The discount factor α is .98 and p, the transition probability from state 1 to state 1, is .9. This numerical example has the same number of cost outcomes as Kaplan's. The parameter values are changed since the monotone condition did not hold for the λ-stopping problem with his parameters.

For our example the cost of continuing, $C(x,1)$, is given by (3) and is $xm_1 + (1-x)m_2 = 2.4x + (1-x)3 = 3-.6x$. The cost of stopping, $C(x,2)$ is $K = 1$. In the λ-stopping formulation we do not add the cost m_1 to K since the convention is to assume that stopping is instantaneous. The transition probabilities are given by equation (5).

We next consider the monotone condition and determine the set of λ-stopping problems such that the monotone condition holds. Given λ, the set B is $\{x: 1 \leq 3-.6x-\lambda+.98\}$ so that B is always of the form $\{x: x \leq c\}$. Therefore we need to determine those states x such that $n(x)$, the next state given an observed cost of 0, is larger than x. These are the states that potentially can cause us to leave the set B once it is entered and thus violate the monotone condition. The situation where the observed cost is 6 does not need to be considered since the next state always has a lower value when the observed cost is 6 than when it is 0. From (5) we have that $n(x) = .54x/(.5+.1x)$, and therefore $n(x) - x$ is decreasing in x. The x such that $n(x) = x$ is .4 so that for $x \geq .4$ $n(x) \leq x$. Our claim is that the monotone condition will hold when B is of the form $\{x: x \leq c\}$ and $c \geq .4$. The proof is that if $x \in B$ then $n(x) \leq n(c)$ since n is increasing, and $n(c) \leq c \in B$ since $c \geq .4$. In terms of λ the requirement is that $1 \leq 3 - .24 - \lambda + .98$ and therefore that $\lambda \leq 2.74$.

We begin step 0.A. of the algorithm by setting $\lambda_0 = 2.5$. This value of λ_0 was simply a reasonable value that we hope will be less than λ^*. The monotone condition holds so that Theorem 1 says that the λ_0-optimal policy is to stop for $x \leq (3+.98-2.5-1.0)/.6 = .8$.

The next step is to calculate V(2.5). This is done by calculating $V^i(2.5)$, the expected cost up to and including period i of the λ-stopping problem, successively for i = 0,1,2.... .

In the zero period we know that x = 1 with probability 1 and $C(1,1) - \lambda = 3-.6-2.5 = -.1$. Therefore, $V^0(2.5) = -.1$. In the first period we know that x = .9 with probability 1 and $C(.9,1) - \lambda = -.04$. When we include the discount factor of .98 we have that $V^1(2.5) = -.1392$.

In the first period we observe a cost of zero with probability .59 = [(.9)(.6) + (.1)(.5)] and a cost of six with probability .41. Therefore in the

second period the $x = .82373$ with probability .59 and $x < .8$ with probability .41.

Consequently $V^2(2.5) = -.1392 + (.59(.00077) + .41(1))\,(.98)^2 = .25783$. In period three $x < .8$ with probability one and $V = V^3(2.5) = .25783 + (.59)(.98)^3 = .81313$. The expected discounted time until stopping in $1 + .98 + .56663 = 2.54663$. Since $V(2.5) > 0$, $\lambda_o < \lambda^*$ and we can proceed to Step 0.B.

The formula for λ_1 in Step 0.B. has a value greater than 2.74, the largest value for which the monotone condition holds. Therefore we will instead set $\lambda_1 = 2.734$, very close to the upper bound, and hope that $V(2.734) \leq 0$. The value of $V(2.734)$ is found to be $-.44225$ and the expected discounted time until stopping is 8.57073. We proceed to Step 1.

A cubic fit is made of on the interval [2.5,2.734] using (11) and $V_A(x) = 0$ for $x = 2.67667$. Therefore $\lambda = .1\,\lambda_B + .9\,\lambda_A = 2.67070$. We solve the 2.67070-stopping problem and $V(2.67070) = .04109$ and the expected discounted time until stopping is 6.67588. A cubic fit is made in the interval [2.67070, 2.734] and $V_A(x) = 0$ for $x = 2.67677$, and $\lambda = 2.67933$. The value of $(1-\alpha)F^{\alpha}_{\pi}(x^o) = 2.67070 + .04109/6.67588$ so that $\lambda^{new} = 2.67686$. We solve the 2.67686-stopping problem and $V(2.67686) = -.00046$ and the expected discounted time until stopping is 6.84113. At this point we apply Proposition 2 with $\lambda_0 = 2.67070$ and $\lambda_1 = 2.67686$. The value of $F^{\alpha}_{\pi_1}(x^o) - F^{\alpha}_{\pi^*}(x^o) \leq \frac{1}{.02}\,\frac{(.00046)(.16525)}{(6.67588)(6.84113)} = .00008$. Since $F^{\alpha}_{\pi_1}(x^o) = \frac{1}{.02}(2.67686 - .00006) = 133.84$ the cost of policy π_1 is within .00005 per cent of the true minimum, $F^{\alpha}_{\pi^*}(x^o)$. Without Proposition 2 we would simply compare the lowest cost policy with the known lower bound. The cost of the 2.67686-stopping problem is by (10) $2.67686 - (.00046/6.84113) = 2.67679$ and the lower bound for λ^* is 2.6707. The percent error is .228.

The policy π_1 is to stop if and only if the probability of being in state 1 is below .50523. This policy could be obtained by Brown's procedure of recursive

sets of rules [5]. At each iteration a new rule is added and there are occasional deletions. The optimal solution for this example problem requires 25 rules. They are:

Rule	Action	Next Rule 0	Next Rule 6
1	C	2	2
2	C	3	4
3	C	5	8
4	C	6	11
5	C	7	14
6	C	9	16
7	C	10	19
8	C	12	19
9	C	13	20
10	C	14	21
11	C	15	21
12	C	16	21
13	C	16	22
14	C	17	23
15	C	18	23
16	C	18	24
17	C	19	24
18	C	19	25
19	C	20	25
20	C	21	25
21	C	22	25
22	C	23	25
23	C	24	25
24	C	25	25
25	S	1	1

This solution is implemented as follows. Assuming that originally we are in state 1 with probability one, we begin with rule 1. The action is to continue. Now suppose that the observed cost is 6. Then we next use rule 2 and our action is to continue. If the observed cost is 0 then we next use rule 3. Again we continue and if the observed cost is 6, then we next use rule 8.

Applying rule 8 our action is to continue and if we observe a cost of 6 then our next rule is 19. Rule 19 says to continue and if the next observed cost is 0 then we use rule 20. Rule 20 says to continue and if the next observed cost is 6 then we use rule 25. Rule 25 says to stop and then use rule 1, etc. These rules have the property that when rule 25 is reached, then the state of the system is less than .50523.

The computations involved in adding a new rule in Brown's algorithm are comparable to the work of performing one iteration of the Regenerative Stopping Algorithm as the main effort for both is ascertaining the current optimal return function or its surrogate. As the number of rules generated to solve the example problem must have been at least 25, the Regenerative Stopping Algorithm does quite well in comparison.

5. THE GENERALIZED KAPLAN MODEL

We consider an n-dimensional version of Kaplan's model where system j, $1 \leq j \leq n$, is a two-state production system as before, and equations (1), (2), and (5) apply to each system j. The probability laws of the n systems are assumed to be independent. The decisions stop and continue apply to all n systems. With the decision continue the total cost is the sum of n different continue costs as described by (3). With the decision stop the total cost is $K + \sum_{j=1}^{n} m_1(j)$ where $m_1(j)$ is the expected cost when system j is in state 1 with probability one. This problem is n-dimensional and the state of the system is $(x(1), \ldots, x(n))$ where $x(j)$ is the probability that system j is in state one. Value iteration, policy iteration, and the Regenerative Stopping Algorithm can be applied to the generalized model, and each will soon fall to the curse of dimensionality. In the case of the Regenerative Stopping Algorithm the difficulty lies in calculating $V(\lambda)$. For example when n reaches 10 and each $x(j)$, $1 \leq j \leq n$, may take on up to 30 values, the problem is computationally intractable. However, the Regenerative Stopping Algorithm has the advantage that

$V(\lambda)$ for a given λ is a number so that it is feasible to estimate $V(\lambda)$ by simulation and this will be carried out. Simulation cannot be applied for value iteration or policy iteration since each iteration requires that an optimal return function be determined.

Before considering a higher dimension problem we apply simulation to the example problem of Section 4. The Regenerative Stopping Algorithm applies except that in Step 1, $\lambda^{new} = \lambda$ since we would not know $(1-\alpha)\ F_{\pi}^{\alpha}(x^o)$ with certainty. The main difference comes after the iterations have been completed and we evaluate the results using Proposition 2.

For each iteration we generate a sample size of 1000. Variations of the variance reduction methods described in Wagner [21] were used, and they resulted in a reduction of variance in the order of 5 to 30 times as compared to the completely independent case.

The results of the simulation are:

Iteration	λ	$V(\lambda)$	Standard Deviation of Error of $V(\lambda)$	$T(\lambda)$	Standard Deviation of Error of $T(\lambda)$
1	2.5	.81313	.00007	2.54663	.00278
2	2.734	-.43986	.00172	8.55863	.02632
3	2.67091	.03893	.00111	6.71135	.02134
4	2.67923	-.01618	.00116	6.90592	.02197

In the algorithm the values of the standard deviations are not used. However, they are used when we apply Proposition 2 which we now describe. We let $\lambda_0 = 2.67091$ and $\lambda_1 = 2.67923$. For λ_1 we set $V(\lambda_1) = -.01618 - 5(.00116) = -.022$. For $T(\lambda_1)$ and $T(\lambda_0)$ we use $6.90592 - 5(.02197)$ and $6.71135 - 5(.02134)$ respectively. Thus in the calculations of $V(\lambda_1)$, $T(\lambda_1)$, $T(\lambda_0)$ we are conservative and are using a value of five standard errors from the estimated mean. In order to calculate $[T(\lambda_1)-T(\lambda_0)]$ we will make an additional simulation of size 100 to calculate this quantity directly. The result here is that $T(\lambda_1)-T(\lambda_0)$ has a mean of .26799 and a standard error of .02391. Thus with very high

probability

$$\left(F^{\alpha}_{\pi_1}(x^o) - F^{\alpha}_{\pi^*}(x^o)\right) \le \frac{1}{.02}\,\frac{(.022)(.38794)}{(6.60465)(6.79607)}$$

$$= .00950$$

Since $F^{\alpha}_{\pi^*}(x^o) \sim 2.679 \times 50$ the percentage of error is, with very high probability, less than or equal to .007 percent of the optimum. In this case, we can calculate the true percentage error for the policy which is optimal for the 2.67923-stopping problem using the non-simulation algorithm. The true percentage of error was .0001.

The n-dimensional example that we will consider has n = 20 and the probability law of each independent system is the same as the example problem we have been considering. For system j the costs 0 and 6 are changed to 0 and 6j with the same probabilities as before so that system 1 is precisely the example problem. We set the cost of stopping, K, equal to 160.

As before, we must consider the monotone condition and determine the set of λ-stopping problems such that the monotone condition holds. Given λ the set B is $\{x(1), \ldots, x(20)\colon K \le \sum_{j=1}^{20} j\left(3 - .6\,x(j)\right) - \lambda + .98K\}$ so that B is of the form $\{x(1), \ldots, X(20)\colon \sum_{j=1}^{20} jx(j) \le c\}$. Therefore we need to determine those vectors x such that n(x), the next vector given that each system observed a cost of 0, has $\sum_{j=1}^{n} jn(x(j)) \ge \sum_{j=1}^{n} jx(j)$. For each system j, $n(x(j)) = .54\,x(j)/(.5 + .1\,x(j))$. Our claim is that the monotone condition will hold if c is $\ge \sum_{j=1}^{20} .4j = 84$, and the proof is as follows.

Let $x \in B$ so that $\sum_{j=1}^{20} jx(j) \le c$. Now let z be a number such that $\sum_{j=1}^{20} jz = \sum_{j=1}^{20} jx(j)$. From the one dimension case in Section 4 we know that $n(z) \le n(c/210)$ since n is increasing and $n(c/210) \le c/210$ when $c \ge 84$, so that

$210\, n(z) \le c$. The proof is complete when we show that $\sum_{j=1}^{20} jn(z) \ge \sum_{j=1}^{20} jn(x(j))$.

We have $\sum_{j=1}^{20} jn(z) = \sum_{j=1}^{20} jn(x(j)) + \sum_{j=1}^{20} j(n(z)) - n(x(j))$. By taking derivatives one can show that n is concave, and the concavity of n implies that $\sum_{j=1}^{20} j(n(z)) - n(x(j)) \ge \sum_{j=1}^{20} jn'(z)[z - x(j)] = 0$ since $\sum_{j=1}^{20} j[z - x(j)] = 0$ by construction, which completes the proof. In terms of λ the requirement is that $K \le 630 - 50.4 - \lambda + .98K$ and $\lambda \le 579.6 - .02K = 576.4$.

The results of the simulation are:

Iteration	λ	$V(\lambda)$	Standard Error of $V(\lambda)$	$T(\lambda)$	Standard Error of $T(\lambda)$
1	525.	124.07	.027	2.6484	.018
2	575.	-132.674	.1588	8.2694	.0189
3	555.219	3.3418	.0638	5.6715	.01148
4	556.735	-5.2245	.0580	5.7820	.01318
5	555.829	-.0829	.0299	5.7036	.00701
6	555.785	.21977	.03006	5.7027	.00712
7	555.842	-.18692	.0297	5.7143	.007429

The iteration 1-4 were based on a sample size of 250 while iterations 5-7 had a sample size of 1000. Iteration 7 was made since $\lambda_1 = 555.829$ could not be used to implement Proposition 2 since $V(\lambda)$ was within 5 standard deviations of 0 and therefore we are not sufficiently confident that $555.829 > \lambda^*$.

We set $\lambda_o = 555.785$ and $\lambda_1 = 555.842$. For Proposition 2 we set $V(\lambda_1) = -.18692 - 5(.0297) = -.3354$, $T(\lambda_0) = 5.7027 - 5(.00712) = 5.6671$, and $T(\lambda_1) = 5.7143 - 5(.00743) = 5.6772$. We make an additional simulation of sample size 400 to calculate $T(\lambda_1) - T(\lambda_0)$. The result is $T(\lambda_1) - T(\lambda_0)$ has a mean of .00447 and a standard error of .002. Thus with a very high probability

$$\left(F^{\alpha}_{\pi_1}(x^o) - F^{\alpha}_{\pi^*}(x^o)\right) \le \frac{1}{.02} \frac{(.3354)\ (.01447)}{(5.6671)\ (5.743)} = .00745\ .$$

Since $F^{\alpha}_{\pi_1}(x^o) \sim 555 \times 50$ the percentage of error is, with very high probability,

less than .00003. This policy is to continue if and only if

$$\sum_{j=1}^{20} jx(j) > \frac{1}{.6}(630 - 555.842 - 3.2) = 118.2633.$$

REFERENCES

[1] Bell, C., "Improved Algorithms for Inventory and Replacement-Stocking Problems," SIAM J. of Applied Math., 18., 1970, 558-566.

[2] Blackwell, D., "Discrete Dynamic Programming," Ann. Math. Statist., 33, 1962, 719-726.

[3] Breiman, L., "Stopping-Rule Problems," Chapter 10 of Applied Combinatorial Mathematics, E. Beckenback, ed., Wiley, New York, 1964.

[4] Brender, D., "A Surveillance Model for Recurrent Events," IBM Watson Research Center Report, Yorktown Heights, New York, 1963. Also summarized in Barlow, R. and F. Proschan, Mathematical Theory of Reliability, Wiley, New York, 1965, 115-116.

[5] Brown, G. D., "Recursive Sets of Rules in Statistical Decision Processes," Statistical Papers in Honor of George W. Schnedecor Ed., T. A. Bancroft, Iowa State University Press, 1972.

[6] Buckman, A. G., Economics of Information and Accounting Theory, Ph.D. dissertation, Graduate School of Management, University of California, Los Angeles, 1972.

[7] Chow, Y., H. Robbins, and D. Siegmund, Great Expectations: The Theory of Optimal Stopping, Houghton Mifflin, Boston, 1971.

[8] Denardo, E., "Contraction Mappings in the Theory Underlying Dynamic Programming," SIAM Review, 9, 1967, 165-177.

[9] Dittman, D. and P. Prakash, "Cost Variance Investigation: Markovian vs. Optimal Control," The Accounting Review, 54, 1979, 358-373.

[10] Dyckman, T., "The Investigation of Cost Variances," J. of Accounting Research, 7, 1969, 215-244.

[11] Kaplan, R., "Optimal Investigation Strategies with Imperfect Information"; J. of Accounting Research, 7, 1969, 32-43.

[12] Kaplan, R., "The Significance and Investigation of Cost Variances: Survey and Extensions," J. of Accounting Research, 13, 1975, 311-337.

[13] Magee, R. P., "A Simulation of Alternative Cost Investigation Models," The Accounting Review, 51, 1976, 529-544.

[14] Miller, B., "Countable State Average Cost Regenerative Stopping Problems," Western Management Science Institute, UCLA Working Paper No. 288, 1979.

[15] Ozan, T., and T. Dyckman, "A Normative Model for Investigating Decisions Involving Multiorigin Costs," J. of Accounting Research, 9, 1971, 88-115.

[16] Pierskalla W., and J. Voelker, "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," Naval Research Logistics Quarterly, 23, 1976, 353-388.

[17] Porteus, E., "Bounds and Transformations for Discounted Finite Markov Decision Chains," Operations Research, 23, 1975, 761-784.

[18] Ross, S., Applied Probability Models with Optimization Applications, Holden-Day, San Francisco, 1970.

[19] Sondik, E. J., "The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs," Operation Research, 26, 1978, 282-304.

[20] Taylor, H., "Optimal Stopping in a Markov Process," Ann. Math. Statist., 39, 1968, 1333-1344.

[21] Wagner, H. M., Principles of Operations Research, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.